



Automatic non-destructive multiple lettuce traits prediction based on DeepLabV3 +

Yu Zhang^{1,4} · Mengliu Wu¹ · Jinsong Li^{1,4} · Si Yang^{1,4} · Lihua Zheng^{1,3,4} · Xinliang Liu^{1,2} · Minjuan Wang^{1,3,4}

Received: 2 June 2022 / Accepted: 7 October 2022 / Published online: 25 October 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

In crop growth management, phenotypic traits are an important basis for judging growth status. Manual measurements are labor-intensive, unstable and time-consuming. We propose an image processing pipeline to estimate multiple lettuce traits (fresh weight, dry weight, plant height, diameter, leaf area) based on the lightweight DeepLabV3 + network. Accurate and rapid segmentation of crops from backgrounds is the basis for phenotypic research. First, we propose to combine DeepLabV3 + and MobilenetV2 to realize a high-precision and fast segmentation of lettuce in complex backgrounds and illuminations. Based on the segmentation results, we extracted the morphological factors and vegetation indices. Random forest (RF), partial least squares regression (PLSR) and support vector machine were applied to predict the multiple lettuce traits and compared for optimal model selection. Results showed that DeepLabv3 + (with Mobilenetv2) has the best segmentation performance with pixel accuracy of 97.520% and 99.821%, mIoU of 88.661% and 98.517%, and segmentation speeds with 0.094 and 0.049 ms per image in dataset D3 and dataset D4. PLSR had the highest accuracy in predicting fresh weight, dry weight, diameter and leaf area, with R^2 of 0.898, 0.899, 0.931 and 0.904, respectively. RF yielded the highest accuracy in predicting plant height, with R^2 of 0.858. We proposed method for estimating phenotypic characteristics of lettuce based on deep learning has excellent performance and important application value for lettuce growth monitoring and yield estimation.

Keywords Machine vision · Plant multiple traits prediction · Machine learning · Shoot biomass

Introduction

Lettuce is the most consumed and cultivated leafy vegetable in the world. It is rich in nutrients such as protein, vitamin E, vitamin C, potassium, calcium and copper [1]. The market demand for lettuce is large and the output continues to increase every year, but its growth period is short and perishable [2]. Real-time dynamic detection of lettuce growth and nutritional status will help producers to manage scientifically and efficiently, reduce resource waste, and improve crop productivity. Plant phenotype is the external expression of plant genes, the result of complex interactions between environmental conditions and genes [3], and is also an important parameter reflecting plant growth and development. Therefore, real-time dynamic monitoring of lettuce phenotypic traits without destructive sampling is of great significance for the realization of precise planting and management of lettuce.

The traditional plant phenotype measurement methods are manual sampling and destructive measurement, which are time-consuming, labor-intensive and greatly affected

✉ Lihua Zheng
zhenglh@cau.edu.cn
✉ Xinliang Liu
liuxinl@btbu.edu.cn
✉ Minjuan Wang
minjuan@cau.edu.cn

¹ College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China

² National Engineering Laboratory for Agri-Product Quality Traceability, Beijing Technology and Business University, Beijing 100048, China

³ Yantai Institute of China Agricultural University, Yantai 264670, China

⁴ Key Laboratory of Smart Agriculture Systems, Ministry of Education, China Agricultural University, Beijing 100083, China

by the subjective factors of the experimenters [3]. Destructive measurements can also cause irreversible damage to plants, making plants unable to continue to grow. The traditional plant phenotype measurement method is inefficient, unable to meet the needs of high-throughput and automated phenotypic analysis, and cannot achieve real-time monitoring. At present, the sensors used in plant phenotyping research mainly include red-green-blue (RGB) cameras [4, 5], thermal cameras [6, 7] fluorescence sensors [8–10], multispectral cameras [6], hyperspectral cameras [11, 12] and lidar devices [13, 14], etc. Compared to other sensors, RGB cameras have the advantage of being affordable and providing high-resolution images. RGB cameras have been widely used in many research areas such as region detection [15, 16], disease monitoring [17, 18] 3D model construction [19–21], etc. Many researchers have used RGB cameras to collect plant information for phenotype research. Han et al. used HSV threshold method to accurately segment walnut seed images, and then quantitatively described walnut seed morphology, color and texture [22]. Zhang et al. realized automatic measurement of fruit spot area, number, size and color difference between fruit spot and epidermis by analyzing the images of pear epidermis samples [23]. The measurement accuracy and precision of this method are comparable to manual methods. Zhang et al. extracted color features from UAV RGB images and established the maize yield estimation model [24]. The range of mean absolute percentage error (MAPE) is from 6.2 to 15.1%. Song et al. designed a high throughput automatic seed test device for maize based on computer vision technology, which realized the extraction and measurement of parameters such as grain thickness, grain length, grain width, quality, length of maize ear and width of maize ear [25].

Accurate segmentation of plants and backgrounds is the primary task for phenotypic research. Compared with the traditional machine vision methods, deep learning can actively learn image features. The robustness and accuracy of deep learning are higher. With the continuous development of deep learning, many deep learning models are applied to the field of leaf segmentation. Ren et al. combined the attention mechanism with the end-to-end neural network to segment the leaves [26]. Xu et al. used Mask R-CNN model for leaf segmentation, and the accuracy of leaf segmentation was significantly improved [27]. Wang et al. fused DeepLabV3 + with U-Net for cucumber leaf segmentation, and the accuracy of leaf segmentation reached 93.27% [28]. Wu et al. used DeepLabV3 + model to segment lettuce leaves, effectively taking into account the segmentation accuracy and speed [29]. Therefore, it is feasible to segment lettuce from complex images by deep learning.

DeepLabV3 + is a classic deep learning semantic segmentation model [30], which is widely used in many research fields such as medicine [31], agriculture [32] and industry

[33]. DeepLabV3 + combines the characteristics of ASPP module and encoder-decoder, which can capture multi-scale information effectively and segment object boundaries more clearly. However, DeepLabV3 + also has many problems such as many parameters, complex structure, and difficulty in deploying to mobile devices and embedded devices. Therefore, this study combines Mobilenetv2 [34] with DeepLabV3 + network to achieve high-precision and fast lettuce leaf segmentation.

The traditional estimation of plant phenotypic parameters is to establish a linear or nonlinear regression model between the relevant parameters and the measured data. In recent years, due to the stronger generalization ability and nonlinear fitting ability of machine learning, many researchers use machine learning to perform regression estimation of plant phenotypic parameters. Pourshamsi et al. used machine learning to estimate forest canopy height [35]. Zhang et al. used machine learning to estimate the wheat leaf area index [36]. Random forest (RF), partial least squares regression (PLSR) and support vector machine (SVM) are widely used machine learning models in agriculture [37–40].

Therefore, this study makes full use of the features of easy acquisition and high resolution of RGB imaging and uses DeepLabV3 + segmentation network and machine learning model to carry out automatic and non-destructive prediction of lettuce multiple traits (fresh weight, dry weight, plant height, diameter and leaf area). The specific goals of this study are: (1) Construct a lightweight DeepLabV3 + semantic segmentation network and evaluate its ability to segment lettuce in complex backgrounds. (2) Extract morphological factors and vegetation indices from the semantic segmentation results, and establish a convenient and high-precision phenotypic trait estimation method for lettuce by comparing RF, PLSR and SVM models.

Materials and methods

Data acquisition

The public dataset used in this experiment comes from the 3rd Greenhouse Autonomous Challenge jointly organized by Wageningen University and Tencent Laboratory [41]. This dataset contains 4 varieties of lettuce from seedling to mature (Satine, Salanova, Aphylion, Lugano) grown in hydroponics. From March 5, 2021 to April 16, 2021, images were taken every other week with Intel RealSense D415, and at the same time, parameters were measured manually (fresh weight, dry weight, plant height, diameter and leaf area). A total of 391 images (1920×1080 pixels), and their multi-trait and biomass data were obtained after screening. The details of the cultivation pattern and the measurement

criterion can be seen in [41]. In this paper, the public dataset is defined as dataset *D1*.

To verify the robustness of the segmentation model, in addition to public datasets, this paper also used self-collected plant factory lettuce images. The experiment was carried out in the plant factory of the School of Information and Electrical Engineering, China Agricultural University, Beijing, and the variety of lettuce was Italian crepe myrtle. The lettuce was grown and photographed in the plant cultivation room. The indoor temperature in the daytime is set to $(26 \pm 1)^\circ\text{C}$, and the indoor temperature at night is set to $(18 \pm 1)^\circ\text{C}$, with LED lighting. The illumination time is maintained at about 18 h each day. The image acquisition device is a SONY SELP1650 digital camera with a focal length of 100 mm and an exposure time of 3 ms. The schematic diagram of the lettuce canopy image acquisition is shown in Fig. 1. To determine the image acquisition range, place the lettuce plants in the planting basket in the center of the white foam board. A 0.250 m-long blue square centered on the planting basket serves as an auxiliary on the white foam board. At the same time, in order to improve the clarity of the image, a white background plate is placed outside the blue square contour. When shooting, the handheld camera is 0.600 m from the top of the crop canopy for data acquisition. 42 seedlings with similar growth potential were randomly selected. Since December 8, 2019, image acquisition has been carried out from 21:00 to 22:00 every day for 30 consecutive days. A total of 1260 images were collected, and

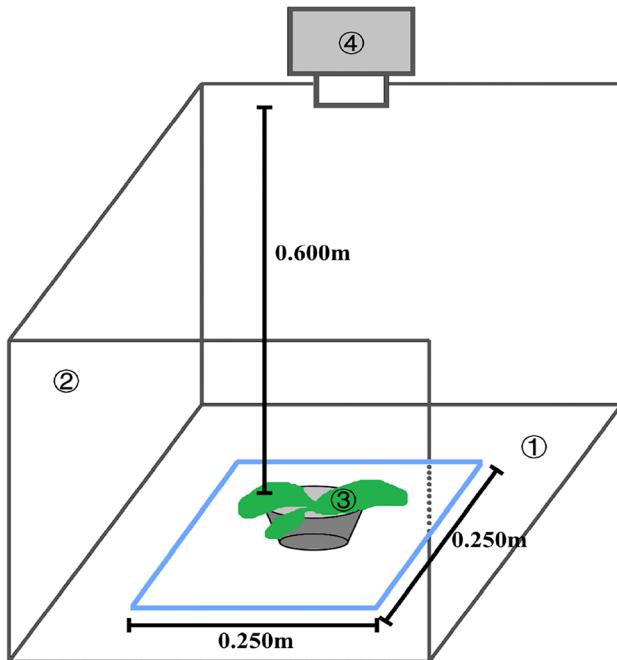


Fig. 1 Schematic diagram of image acquisition ① White foam board ② White background plate ③ The lettuce plants in the planting basket ④ The RGB camera

the resolution of each image was 2576×3872 pixels. In this paper, the self-collected dataset is defined as dataset *D2*.

Data annotation and augmentation

Image sizes that are too large will cause GPU memory exhaustion. Downscaling the image not only saves computational costs but also serves as the basis for low-pixel-resolution vision sensor research. In this paper, image cropping was performed on the public dataset *D1* and the self-collected dataset *D2*. The specific operations of cropping are as follows.

The cropping range of dataset *D1* is 190 to 980 in y and 510 to 1580 in x, creating a 1070×790 pixels image. This cropping range crops out most of the useless information and avoids cutting off the lettuce image. The cropped dataset *D1* is shown in Fig. 2a. Centered on the image center, the dataset *D2* image size is cropped from 2576×3872 pixels to 512×512 pixels. The cropped dataset *D2* is shown in Fig. 2b.

In the field of deep learning, annotation tools such as LabelMe, LabelImg, RectLabel, VOTT, etc. are usually used to manually annotate dataset images. However, due to the many curls and folds on the edges of the lettuce leaves, the labeling accuracy of LabelMe, LabelImg and other tools is not satisfactory. To improve the labeling accuracy, Photoshop software (Adobe, San Jose, USA) was used to label the lettuce canopy image. After labeling, a 24-bit PNG image is generated, with 0 representing the backgrounds (black) and 255 representing the lettuce (white). The original image and annotation results are shown in Fig. 3.

The deep learning training process requires a large amount of data for training to enhance model robustness and reduce overfitting. Augmentor is a 2D image random enhancement library in python [42], which can perform data enhancement on the original image and the labeled image synchronously. Simultaneously upscale the original image and the marker image by randomly using the scaling, rotation, distortion, clipping, and brightness changes included in Augmentor. Database *D1* is recorded as database *D3* after data augmentation and database *D2* is recorded as database *D4* after data augmentation.

Network architecture of DeepLabV3 +

The DeeplabV3 + model introduces an encoder-decoder structure, which can combine low-dimensional features with high-dimensional features to obtain rich spatial information and semantic information, and improve the accuracy of boundary segmentation. The image data is input into the encoder, and the low-level features obtained by deep convolution neural network (DCNN) are output to the decoder. Then, high-level information is obtained by using parallel

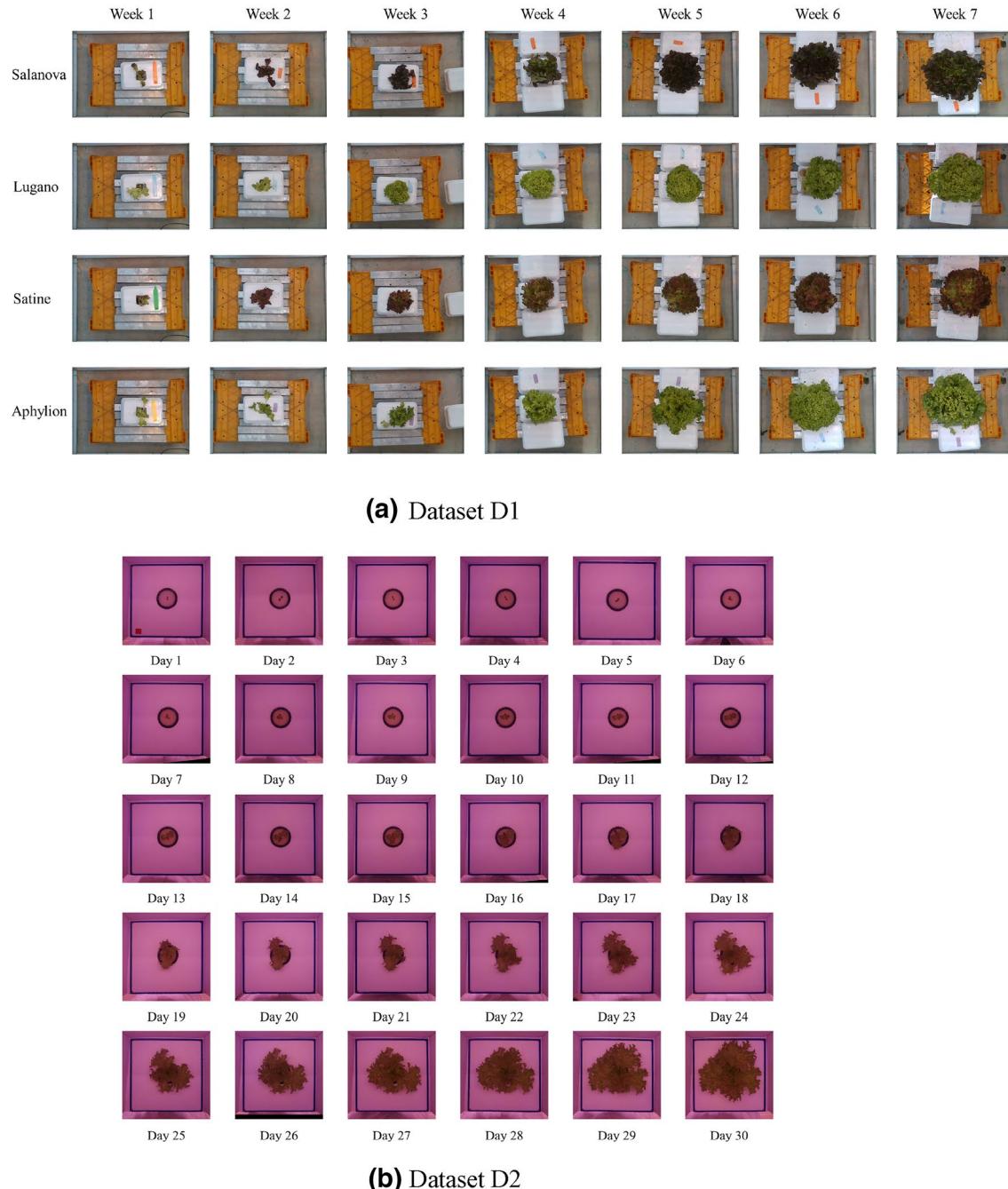


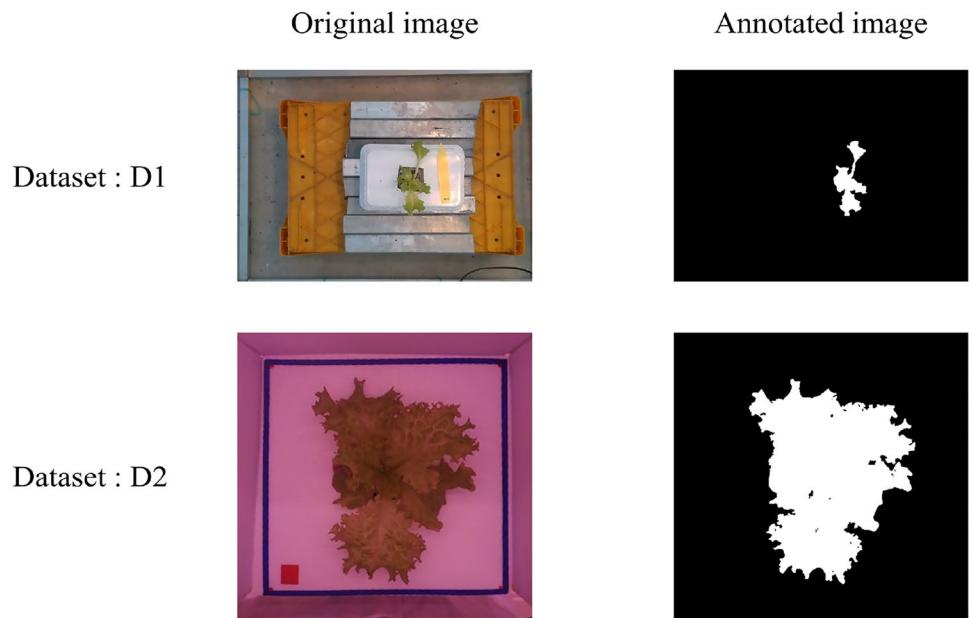
Fig. 2 Schematic diagram of the cropped dataset (a) Dataset D1, the public dataset (b) Dataset D2, the self-collected dataset

atrous convolutions at different rates in the Atrous Spatial Pyramid Pooling (ASPP) module. The ASPP module concatenates the feature maps to generate combined features, and the combined features are fused and reduced dimension through 1×1 convolution. The decoder fuses the low-level features extracted by the backbone network with the high-level features that have been bilinearly upsampled by 4 times and then uses 3×3 convolution to further fuse the features. After the two features are fused, bilinear interpolation is

finally used to obtain a segmentation prediction with the same size as the original image.

This paper replaces the backbone feature extraction network Xception [43] of DeeplabV3 + with Mobilenetv2. Through this operation, the DeeplabV3 + is lightweighted. Mobilenetv2 combines Depthwise Convolution with Inverted Residual and Linear Bottleneck, which effectively reduces model computation while ensuring model accuracy.

Fig. 3 Schematic diagram of dataset annotation



The modified DeeplabV3+ model structure is shown in Fig. 4.

Training with DeepLabV3 +

The experimental computer operating system is Ubuntu16.04, the processor is 6-core AMD Ryzen5 3600, 16 GB memory, the video card is NVIDIA GeForce RTX 3060, and the video random access memory is 12 GB. The deep learning segmentation model is deployed based on pytorch1.8.2 [44], with Python3.8, Cuda10.2 computing

architecture and GPU acceleration library Cudnn8005 installed.

In this study, the transfer learning method was used to initialize the network parameters, and the pre-training weights were derived from the PASCAL VOC data (PASCAL Visual Object Classes) pre-training model. The initial learning rate of the model is 5×10^{-4} , the optimization algorithm is Adam, the loss function is the Cross Entropy Loss, the decay rate is 0.1, and the batch size is set to 8. The number of training iterations is 400. In order to maintain the initial weight of training and speed up the training, some neural networks are frozen in

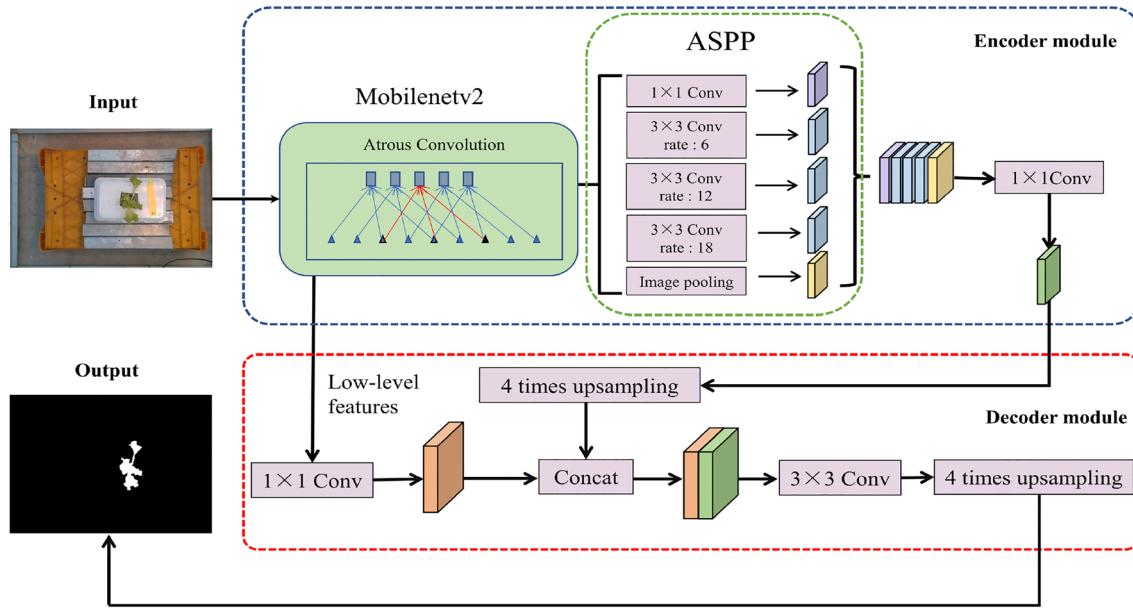


Fig. 4 DeepLabV3+ - Mobilenetv2 network architecture

the first 100 epochs, and all neural networks are unfrozen for training in the last 300 epochs, and the learning rate is adjusted to 5×10^{-5} .

In order to verify the results of DeeplabV3 + network segmentation, this paper selected three traditional segmentation methods (Excess Green (ExG), HSV, GrabCut) [45–47] and two widely used semantic segmentation networks (U-Net, PSPNet) [48, 49] to compare the segmentation results.

Segmentation evaluation index

In order to quantify and compare the segmentation results of lettuce images by different methods, Pixel Accuracy (PA, %), Intersection over Union (IoU, %) and the mean Intersection over Union (mIoU, %) were used to evaluate the segmentation effect.

PA is the percentage of correctly classified pixel number to total pixel number in the testing set, as shown in Eq. 1. IoU is used to describe the degree of coincidence between the prediction box and the label box (Eq. 2). The numerator of the formula is the intersection of the prediction box and the label box, and the denominator of the formula is the union of the prediction box and the label box. mIoU is the average value of the IoU of all classes (Eq. 3). The average segmentation time is used to represent the speed of image segmentation by different methods, and the formula is shown in Eq. 4.

$$PA = \frac{\sum_{i=0}^N p_{ii}}{\sum_{i=0}^N \sum_{j=0}^N p_{ij}} \times 100\% \quad (1)$$

$$IoU = \frac{TP}{TP + FP + FN} \times 100\% \quad (2)$$

$$mIoU = \frac{\sum_{i=0}^N IoU}{N + 1} \times 100\% \quad (3)$$

$$T = \frac{\sum_{i=0}^M t_i}{M + 1} \quad (4)$$

While $\sum_{i=0}^N p_{ii}$ represents the total number of correctly predicted pixels, and $\sum_{i=0}^N \sum_{j=0}^N p_{ij}$ represents the total number of pixels. True Positive (TP) represents the positive samples that are predicted to be positive by the model. False Positive (FP) represents the negative samples that are predicted to be positive by the model. False Negative (FN) represents the positive samples that are predicted to be negative by the model. $N + 1$ represents the sample Number of categories (including background categories). t_i represents the segmentation time of each image, $M + 1$ represents the total number of images.

Index extraction

Morphological factor extraction

The RGB canopy image of lettuce contains a lot of morphological information, which can effectively reflect the growth status of lettuce. Therefore, this paper proposes 8 morphological factors to estimate lettuce morphological parameters and biomass. The definitions and descriptions of the 8 morphological factors extracted from the segmented images are shown in Table 1. The area₂, area₃ and area₄ are all calculated by cv2 library in python. d₂ is obtained from area₂ according to Eq. 5, d₄ is obtained from area₁ according to Eq. 5. Some visualization results are shown in Fig. 5.

$$d = 2\sqrt{\frac{S}{\pi}} \quad (5)$$

Vegetation indices extraction

According to the literature and related research, this paper selected 17 common vegetation indices, the specific content is shown in Table 2. The segmented binary image (Fig. 6a) was mapped to the original image (Fig. 6b) to obtain the

Table 1 Description of morphological factors from the segmented images

| Type | Morphological factor | Description | Number |
|----------|----------------------|---|--------|
| Area | area ₁ | The number of pixels with a value of 255 | I |
| | area ₂ | Contour area of image convex hull after segmentation | II |
| | area ₃ | The minimum circumscribed circle area of the image after segmentation | III |
| | area ₄ | Maximum contour area of the segmented image | IV |
| Diameter | d ₁ | Euclidean distance of the farthest lettuce pixel in the image | V |
| | d ₂ | The diameter of area ₂ (calculated by Eq. 5) | VI |
| | d ₃ | The minimum circumscribed circle diameter of the image after segmentation | VII |
| | d ₄ | The diameter of area ₁ (calculated by Eq. 5) | VIII |

Fig. 5 Schematic diagram of morphological factors of lettuce canopy image after segmentation (a) The number of pixels with a value of 255 (b) Contour area of image convex hull after segmentation (c) The minimum circumscribed circle area of the image after segmentation (d) Maximum contour area of the segmented image (e) Euclidean distance of the farthest lettuce pixel in the image (f) The minimum circumscribed circle diameter of the image after segmentation

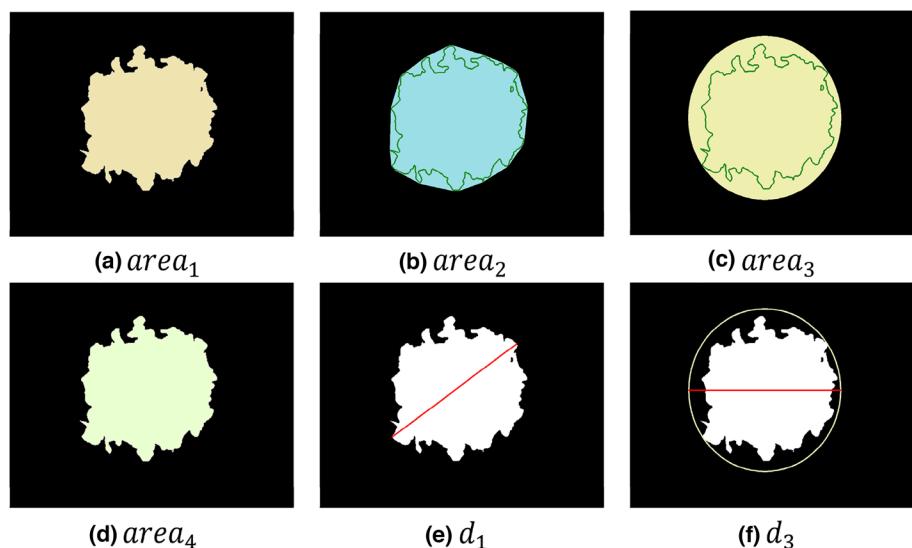


Table 2 Description of vegetation indices from the segmented images

| Image index | Definition | References |
|---|---|------------|
| Red (R), Green (G), Blue (B) | raw value of each band | – |
| Nomalized Red (r), Green (g), Blue (b) | $r=R/(R+G+B)$ $g=G/(R+G+B)$ $b=B/(R+G+B)$ | – |
| Gray grayscale value (Gray) | $Gray=0.3R+0.59G+0.11B$ | [50] |
| Green blue ratio index (GBRI) | $GBRI=G/B$ | [51] |
| Red blue ratio index (RBRI) | $RBRI=R/B$ | [51] |
| Green red ratio index (GRRI) | $GRRI=G/R$ | [51] |
| Excess green index (ExG) | $ExG=2G-R-B$ | [52] |
| Excess red index (ExR) | $ExR=1.3 R-G$ | [52] |
| Excess green minus excess red (ExGR) | $ExGR=ExG-ExR$ | [52] |
| Normalized green-red difference index (NGRDI) | $NGRDI=(G-B)/(G+B)$ | [53] |
| Normalized difference index (NDI) | $NDI=(r-g)/(r+g+0.01)$ | [54] |
| Green-red vegetation index (GRVI) | $GRVI=(G-R)/(G+R)$ | [55] |
| Visible atmospherically resistance index (VARI) | $VARI=(G-R)/(G+R-B)$ | [56] |

color lettuce canopy image (Fig. 6c), and 17 vegetation indices were extracted based on the color lettuce canopy image.

Statistical analysis and modelling

In this paper, 8 morphological factors and 17 vegetation indices are extracted from the segmented images. The correlation between it and the predicted multi-trait (shown in Table 3) was calculated using python. Six highly correlated characteristic variables were selected as modeling factors to construct the lettuce multi-trait estimation model. Three

widely used machine learning regression methods [random forest (RF), partial least squares regression (PLSR) and support vector machine (SVM)] were used to build the estimation model. RF is based on bootstrap sampling, which randomly selects training samples from samples with replacement during training. The sampled samples are trained to build a decision tree. Finally, the average of the output results of the decision tree is used as the regression output of the random forest. PLSR integrates the advantages of principal component analysis (PCA), canonical correlation analysis (CCA), and linear regression analysis. PLSR

Fig. 6 Schematic diagram of the acquisition of color lettuce canopy images (a) the segmented binary image (b) the original image (c) the color lettuce canopy image

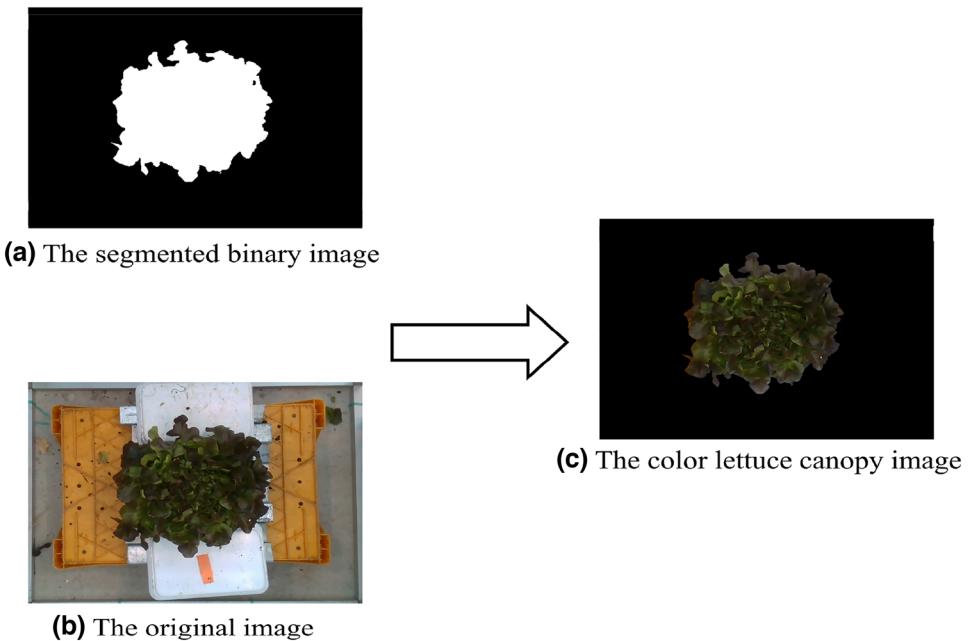


Table 3 Lettuce multiple traits

| Name | Abbreviation | Unit |
|---------------|--------------|-----------------|
| Fresh weight | <i>fw</i> | g |
| Dry weight | <i>dw</i> | g |
| Height | <i>h</i> | cm |
| True diameter | <i>td</i> | cm |
| True area | <i>ta</i> | cm ² |

can effectively resolve multiple correlations between variables. SVM is a widely used machine learning method, which can be divided into linear vector machines and nonlinear vector machines. SVM assumes the existence of a hyperplane that can perform binary classification on the training data. The hyperplane that clearly splits the training data and has the largest distance from the data is considered the optimal hyperplane.

In this study, 391 sets of data (*dataset D2*) were used for model construction. The dataset *D2* was randomly divided into training set and validation set with a ratio of 4:1 before model building. Based on the coefficient of determination [R^2 (Eq. 6)], the optimal hyperparameters for each model were obtained by applying ten-fold cross-validation on the training set. Apply the test dataset to the optimized model and use the coefficient of determination (R^2), Root Mean Square Error ($RMSE$), and Relative Root Mean Square Error

($RRMSE$, %) to evaluate the model accuracy. The specific calculation equation is shown in Eqs. 6, 7 and 8.

$$R^2 = 1 - \frac{\sum_{i=0}^{n+1} (y_i - \hat{y}_i)^2}{\sum_{i=0}^{n+1} (y_i - \bar{y})^2} \quad (6)$$

$$RMSE = \sqrt{\frac{\sum_{i=0}^{n+1} (y_i - \hat{y}_i)^2}{n + 1}} \quad (7)$$

$$RRMSE = \frac{RMSE}{\bar{y}} \times 100\% \quad (8)$$

y_i represents the i sample value; \hat{y}_i represents the model prediction value of the i sample; \bar{y} represents the sample average value; $n + 1$ represents the total number of samples.

Results and discussion

Analysis of performance results of lettuce canopy segmentation model

Table 4 shows the segmentation performance of different segmentation methods on the lettuce canopy in the test set. It

Table 4 The average performance of different models

| Dataset | Model | Backbone | PA % | IoU % | | MIoU % |
|---------|-------------|-------------|---------------|---------------|---------------|---------------|
| | | | | Lettuce | Background | |
| D3 | DeepLabV3 + | Mobilenetv2 | 97.520 | 80.077 | 97.246 | 88.661 |
| | | Xception-65 | 97.461 | 79.663 | 97.180 | 88.421 |
| | U-Net | — | 97.455 | 79.625 | 97.174 | 88.400 |
| | | Mobilenetv2 | 96.985 | 76.409 | 96.659 | 86.534 |
| | PSPNet | ResNet101 | 96.774 | 74.905 | 96.430 | 85.668 |
| | | E × G | — | 88.578 | 30.006 | 58.998 |
| | HSV | — | 65.364 | 6.724 | 64.477 | 35.601 |
| | | GrabCut | — | 96.560 | 73.026 | 84.616 |
| | D4 | DeepLabV3 + | Mobilenetv2 | 99.821 | 97.224 | 99.809 |
| | | | Xception-65 | 99.799 | 96.918 | 98.352 |
| | | U-Net | — | 99.786 | 96.707 | 98.239 |
| | | | Mobilenetv2 | 99.355 | 90.461 | 94.887 |
| | | PSPNet | ResNet101 | 99.493 | 92.383 | 99.460 |
| | | | E × G | — | 60.709 | 96.775 |
| | | HSV | — | 98.564 | 78.610 | 88.547 |
| | | | GrabCut | — | 98.568 | 82.275 |
| | | | | | | 98.466 |

The best performing models for each variable are in bold

can be seen that the traditional color threshold segmentation is generally not as effective as deep learning segmentation.

Comparing three traditional segmentation methods, GrabCut has better segmentation effects on the two datasets. However, the GrabCut algorithm is semi-automatic segmentation, which requires manual selection of the foreground area as a reference, so it is difficult to achieve automatic processing of large batches of data. The HSV method performs great on the D4 dataset and performs the worst on the D3 dataset. It can be seen from the data introduction in 2.1 that the D3 dataset contains four lettuce varieties, and the colors of the lettuce leaves are also different. The HSV algorithm is divided according to the color threshold, and it is difficult to process the lettuce images of many varieties of lettuce leaves with different colors at the same time. Comparing three deep semantic segmentation networks, the evaluation indicators obtained by the DeepLabV3 + model are slightly better than the U-Net model and

the PSPNet model. The DeepLabV3 + model which backbone is Mobilenetv2 has the best segmentation effect.

Table 5 shows the prediction time per image of different deep learning models. The DeepLabv3 + model prediction time is significantly shorter than the rest of the segmentation models. Compared with the DeepLabv3 + (Xception-65), the prediction time of the DeepLabv3 + (Mobilenetv2) is reduced by 0.034 fps and 0.032 fps, respectively.

To sum up, the DeepLabv3 + (Mobilenetv2) has the best segmentation effect and segmentation speed, and is suitable for lettuce canopy segmentation in different varieties, different environments, and different growth periods.

The segmentation effects of different deep semantic segmentation networks on the D3 dataset and D4 dataset are shown in Figs. 7 and 8. Figures 7e, f, 8e and f are the segmentation effect of the PSPNet model. The PSPNet model cannot accurately segment the thinner petiole part, but can only segment the larger petiole part

Table 5 Predicted time in different models (fps)

| Dataset | DeepLabv3 + (Mobile-netv2) | DeepLabv3 + (Xception-65) | U-Net | PSPNet (Mobilenetv2) | PSPNet (ResNet101) |
|---------|----------------------------|---------------------------|-------|----------------------|--------------------|
| D3 | 0.094 | 0.128 | 0.167 | 0.234 | 0.120 |
| D4 | 0.049 | 0.081 | 0.253 | 0.198 | 0.075 |

Fig. 7 The segmentation effect of different segmentation models on the D3 dataset. The red box frame selected part is the thinner petiole (a) the ground truth (b) the segmentation result of DeepLabV3+ (Mobilenetv2) (c) the segmentation result of DeepLabV3+ (Xception-65) (d) the segmentation result of U-Net (e) the segmentation result of PSPNet (Mobilenetv2) (f) the segmentation result of PSPNet (ResNet101) (Color figure online)

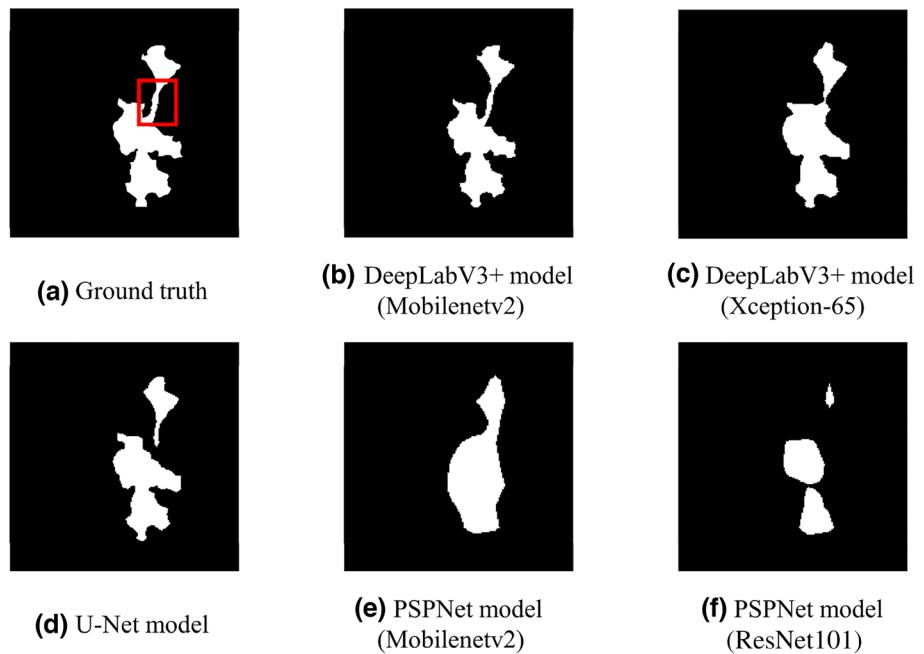


Fig. 8 Segmentation effect of different segmentation models on the D4 dataset. The part selected by the red box is the blade clearance (a) the ground truth (b) the segmentation result of DeepLabV3+ (Mobilenetv2) (c) the segmentation result of DeepLabV3+ (Xception-65) (d) the segmentation result of U-Net (e) the segmentation result of PSPNet (Mobilenetv2) (f) the segmentation result of PSPNet (ResNet101) (Color figure online)

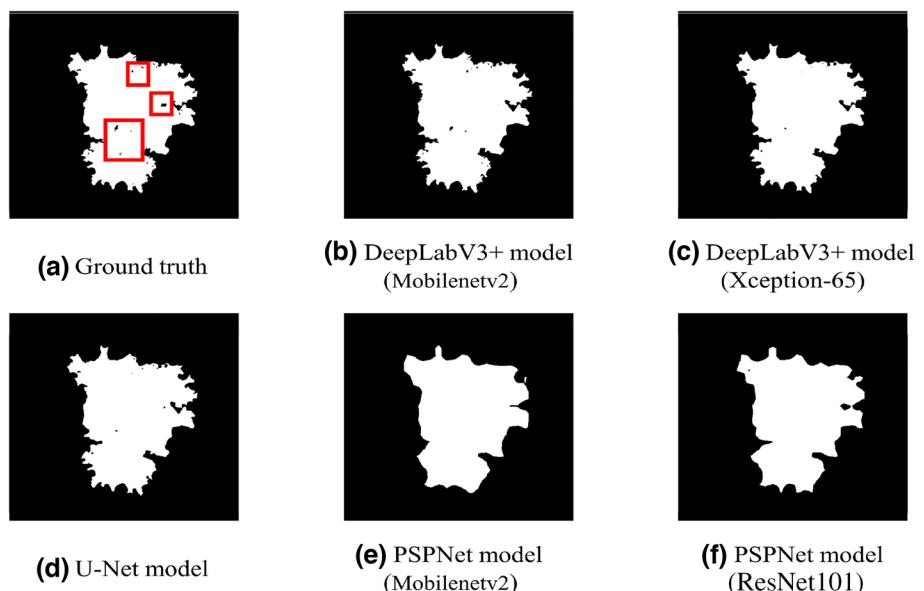


Table 6 Accuracy comparison of leaf semantic segmentation

| Research object | Model | Backbone | PA % | MIoU % | References |
|---------------------------|------------|-------------|-------|--------|------------|
| Lettuce-D3 | DeepLabV3+ | Mobilenetv2 | 97.52 | 88.66 | Ours |
| Lettuce-D4 | DeepLabV3+ | Mobilenetv2 | 99.82 | 98.52 | Ours |
| Lettuce | DeepLabV3+ | ResNet-101 | 99.24 | 83.26 | [29] |
| Lettuce | U-Net | — | | 81.58 | [57] |
| Lettuce | U-Net | — | 99.82 | 87.13 | [58] |
| Coffee, soybean and Wheat | U-Net | — | | 90.23 | [59] |
| Cucumber | MSF-CNNs | — | 92.38 | 91.36 | [60] |

(Fig. 7f). The leaf gap is also wrongly segmented into leaves (Fig. 8e). It can be seen from Figs. 7d and 8d that the U-Net model can segment clearer leaf edges, but there is still a large mis-segmentation phenomenon in the thinner petiole part and the gap between leaves. It can be seen from Figs. 7b, c, 8b and c that the leaf edge segmented by DeepLabV3 + model is clear, and the petiole and leaf gap can be well segmented. Comparing Fig. 7b and c, it can be found that the segmentation of DeepLabV3 + model with Mobilenetv2 is more detailed and the petiole segment is more clearly segmented than the DeepLabV3 + model with Xception-65. The leaf clearance segmentation effect shown in Fig. 8b is better than that in Fig. 8c. This result is also consistent with the evaluation indicators. In summary, the DeepLabV3 + (Mobilenetv2) has better segmentation performance for lettuce of different maturity in different varieties and different environments.

In order to prove the accuracy of the semantic segmentation network proposed in this paper for leaf segmentation, this study summarizes similar articles, and the results are shown in Table 6. The studies shown in Table 6 are all based on the deep learning network to segment the leaves. The research objects involved lettuce, coffee leaves, soybean leaves, wheat leaves and cucumber leaves. It can be seen from Table 6 that the DeepLabV3 + (Mobilenetv2) network proposed in this paper has higher accuracy, indicating that the proposed method can accurately and effectively segment lettuce.

The selection of modeling factors

Correlation analysis of morphological factors and vegetation indices with lettuce phenotypic parameters to be predicted was performed in turn, and the results were shown in Figs. 9 and 10, respectively. It can be seen from

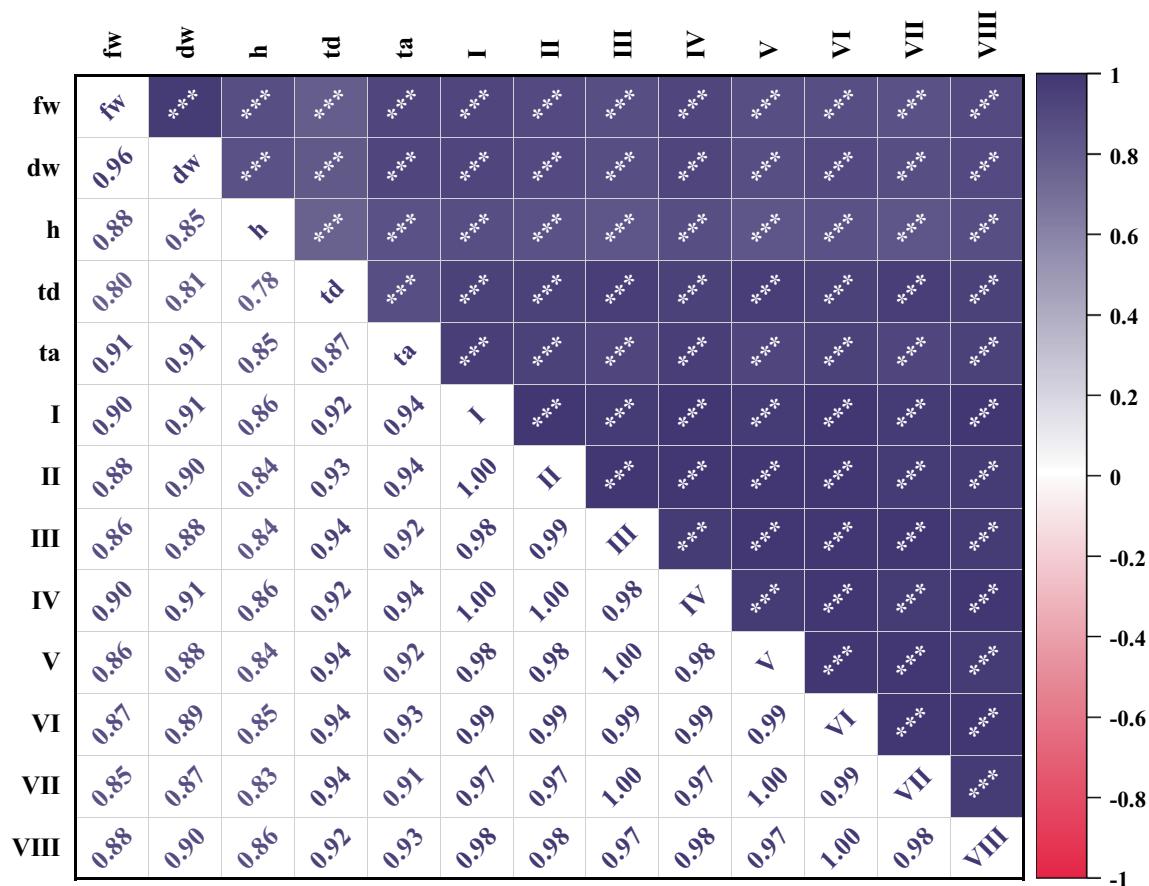


Fig. 9 The correlation analysis of morphological factors. * $p \leq 0.05$ ** $p \leq 0.01$ *** $p \leq 0.001$

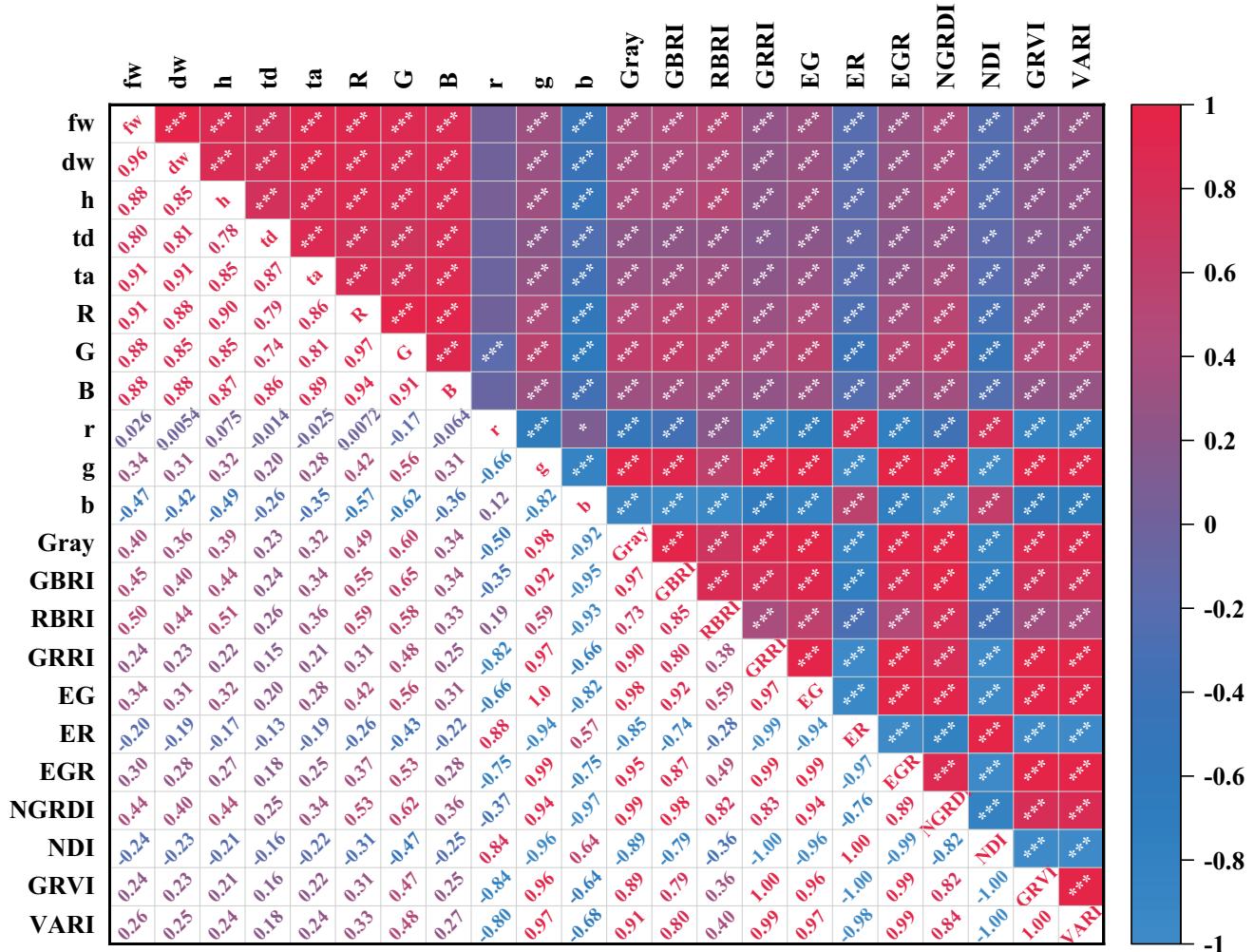


Fig. 10 The correlation analysis of vegetation indices. * $p \leq 0.05$ ** $p \leq 0.01$ *** $p \leq 0.001$

Table 7 Modeling factors

| Phenotypic parameter | Modeling factors |
|----------------------|---------------------|
| <i>fw</i> | I R IV II VII B |
| <i>dw</i> | IV I II VIII VI R |
| <i>h</i> | R B VIII IV I VI |
| <i>td</i> | V III VII II I |
| <i>ta</i> | IV I II VIII VI III |

the Fig. 9 that all the eight morphological factors and the multi-trait were all highly significantly positively correlated ($P \leq 0.001$), and most of them showed a significant linear correlation ($0.5 < |r| \leq 0.8$). However, the correlations of different vegetation indices are quite different. Some factors are positively correlated, some are negatively correlated, some are strongly correlated (such as R, G, B), and some are weakly correlated (such as r).

Among the numerous vegetation indices, R, G, B shows a strong correlation.

The six most relevant factors were selected as the modeling factors to construct the biomass estimation model, as shown in Table 7.

Accuracy verification of biomass estimation model

The six factors with the highest correlation of each phenotypic parameter were used as independent variables, and the estimation models were constructed using RF, PLSR and SVM respectively. Table 8 shows the accuracy evaluation indicators of each model. For *fw*, *dw*, *td* and *ta*, the model with the best training performance is RF, with R^2 of 0.978, 0.967, 0.978, 0.978, respectively. The model with the highest prediction accuracy is PLSR, with R^2 of 0.898, 0.899, 0.931, 0.904, respectively. For *h*, the best models for training and testing are both RF, and R^2 is 0.976 and 0.858, respectively. It can be seen from Table 8 that RF performs best on the

Table 8 Results of biophysical variables retrieval with three models

| Variable | Model | Cross-Validation | | | Test | | |
|------------------------------|-------|------------------|----------------|--------------|----------------|----------------|---------------|
| | | R ² | RMSE | RRMSE(%) | R ² | RMSE | RRMSE(%) |
| <i>fw</i> (g) | RF | 0.978 | 16.068 | 3.625 | 0.862 | 41.471 | 9.063 |
| | PLSR | 0.893 | 35.360 | 7.977 | 0.898 | 35.611 | 7.782 |
| | SVM | 0.854 | 41.216 | 9.298 | 0.851 | 43.055 | 9.409 |
| <i>dw</i> (g) | RF | 0.967 | 0.863 | 2.705 | 0.838 | 1.847 | 9.228 |
| | PLSR | 0.873 | 1.699 | 5.325 | 0.899 | 1.459 | 7.292 |
| | SVM | 0.792 | 2.173 | 6.810 | 0.847 | 1.798 | 8.986 |
| <i>h</i> (cm) | RF | 0.976 | 0.727 | 3.511 | 0.858 | 1.991 | 10.007 |
| | PLSR | 0.827 | 1.936 | 9.351 | 0.842 | 2.100 | 10.551 |
| | SVM | 0.735 | 2.356 | 11.838 | 0.713 | 2.790 | 13.812 |
| <i>td</i> (cm) | RF | 0.978 | 0.889 | 2.631 | 0.908 | 2.105 | 6.972 |
| | PLSR | 0.877 | 2.118 | 6.266 | 0.931 | 1.821 | 6.029 |
| | SVM | 0.824 | 2.536 | 7.503 | 0.888 | 2.328 | 7.707 |
| <i>ta</i> (cm ²) | RF | 0.978 | 229.056 | 3.360 | 0.862 | 549.699 | 8.729 |
| | PLSR | 0.892 | 504.821 | 7.404 | 0.904 | 458.185 | 7.276 |
| | SVM | 0.878 | 536.678 | 7.872 | 0.890 | 491.570 | 7.806 |

The best performing models for each variable are in bold

training set, and the R^2 of five estimations are all higher than 0.960. However, the performance of the model on the test set is quite different from the training set. This shows that the model has poor stability. PLSR training accuracy is slightly lower than RF, but the test accuracy is generally higher than RF model. Moreover, the training accuracy of PLSR model is not significantly different from the test accuracy, indicating that the model is relatively stable. SVM training accuracy and test accuracy are lower than RF and PLSR, but the model stability is stronger. Based on the analysis of the modeling set and validation set of the three models, it can be seen that the model constructed by PLSR has better prediction effect and stronger stability.

The test results of the three models are shown in Fig. 11. The first column is the fitting result based on the RF model, the second column is the fitting result based on the PLSR model, and the third column is the fitting result based on the SVM model. The first row of Fig. 9 is the prediction result of fresh weight of lettuce, the second row is the prediction result of dry weight of lettuce, the third row is the prediction result of plant height of lettuce, the fourth row is the prediction result of lettuce diameter, and the fifth row is the prediction result of leaf area of lettuce.

Conclusion

To replace traditional destructive crop phenotyping, this paper proposes a low-cost, non-contact phenotyping method for lettuce based on an RGB camera. In this study, lettuce canopy images of different varieties, different environments,

and different growth periods were collected by RGB cameras.

Precise segmentation of lettuce from complex backgrounds is the basis for phenotypic parameter extraction research. In this study, DeepLabV3+ was combined with Mobilenetv2 to construct a lightweight lettuce canopy image semantic segmentation network. The segmentation accuracy and segmentation efficiency of this network are better than the original DeepLabV3+ network. On the dataset D3, PA = 97.520%, mIoU = 88.661, and the segmentation speed is 0.094 fps. On the dataset D4, PA = 99.821, mIoU = 98.517, and the segmentation speed is 0.049 fps. The DeepLabV3+ (with Mobilenetv2) can clearly segment the curled part of the lettuce leaf from the thinner petiole part to obtain an accurate lettuce canopy image.

Based on the segmentation results, morphological factors and vegetation indices were extracted. Three regression models (RF, PLSR, SVM) were used to estimate lettuce phenotypic traits (fresh weight, dry weight, plant height, diameter, leaf area). PLSR has high prediction accuracy and stability. For fresh weight, dry weight, diameter, and area, the R^2 of PLSR training and testing were all higher than 0.840, indicating that PLSR has advantages in predicting the above phenotypic parameters. The fitting effect of the RF model for plant height is better than the other two models, with R^2 of 0.858. The prediction effect of plant height is slightly worse than that of other phenotypic parameters. In subsequent studies, image depth information or point cloud data can be added to assist in the fitting of plant height.

In summary, based on RGB images, this paper proposes a non-destructive measurement method for lettuce phenotypic parameters with high accuracy. The method can be

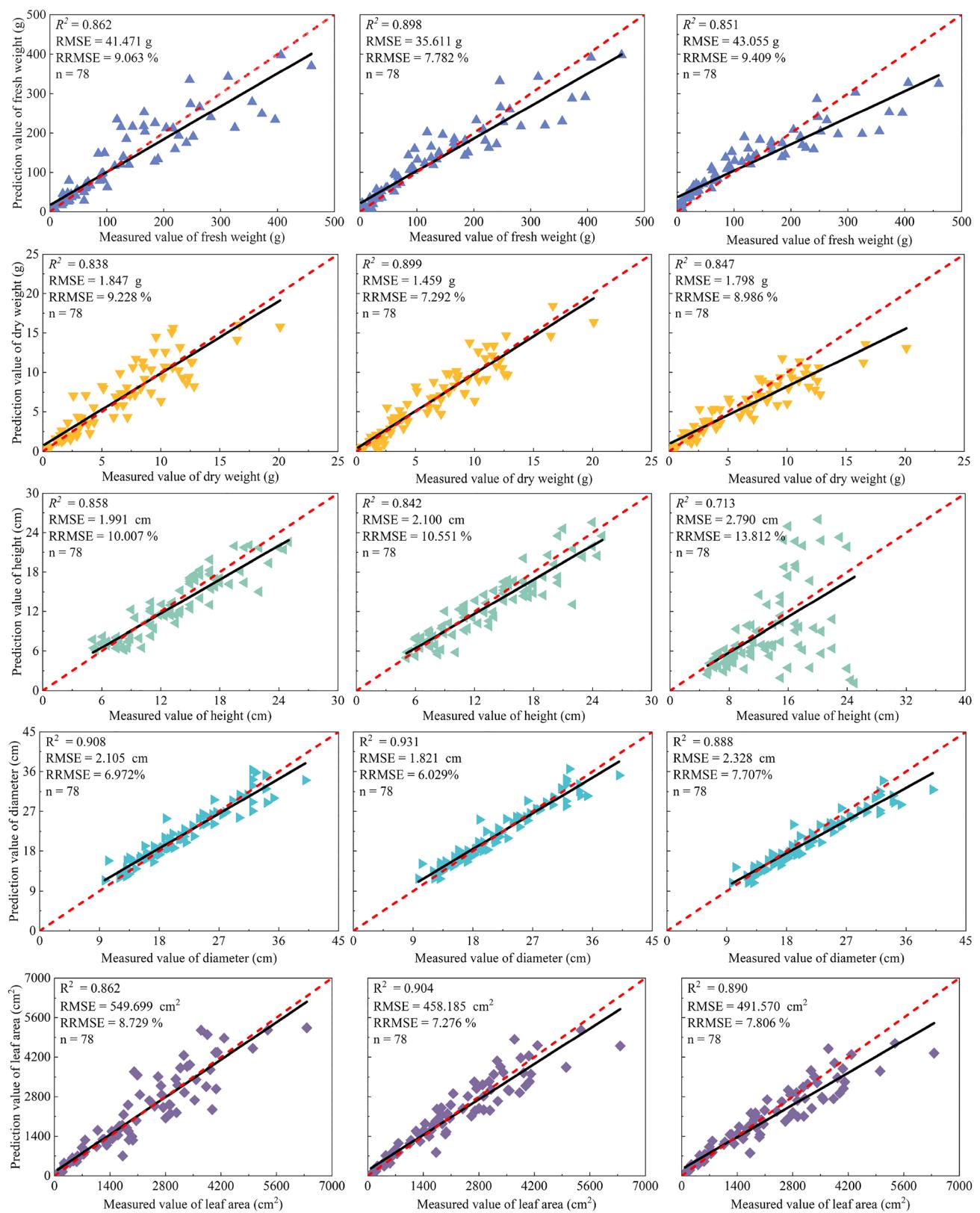


Fig. 11 The fitting results of the predicted and measured values of the sample phenotypic traits obtained by the three methods. The first column is the fitting result based on the RF model, the second column is

the fitting result based on the PLSR model, and the third column is the fitting result based on the SVM model. The black solid lines represent the fitting lines, and the red dotted lines represent the 1:1 lines

used for real-time dynamic management in the growth of lettuce, which is helpful for the management and production of lettuce.

Acknowledgements This work was supported in part by National Natural Science Foundation of China (32201654), Natural Science Foundation of Shandong Province (Grant No. ZR2021MC021) and the Open Project Program of National Engineering Laboratory for Agri-product Quality Traceability (Grant No. AQT-2021-YB6), Beijing Technology and Business University (BTBU).

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- I. Medina-Lozano, J.R. Bertolín, A. Díaz, Nutritional value of commercial and traditional lettuce (*Lactuca sativa* L.) and wild relatives: vitamin C and anthocyanin content. *Food Chem.* **359**, 129864 (2021). <https://doi.org/10.1016/j.foodchem.2021.129864>
- U. De Corato, Improving the shelf-life and quality of fresh and minimally-processed fruits and vegetables for a modern food industry: a comprehensive critical review from the traditional technologies into the most promising advancements. *Crit. Rev. Food Sci. Nutr.* **60**(6), 940–975 (2020). <https://doi.org/10.1080/10408398.2018.1553025>
- X. Jin, P.J. Zarco-Tejada, U. Schmidhalter, M.P. Reynolds, M.J. Hawkesford, R.K. Varshney et al., High-throughput estimation of crop traits: a review of ground and aerial phenotyping platforms. *IEEE Geosci. Remote Sensing Mag.* **9**(1), 200–231 (2020). <https://doi.org/10.1109/mgrs.2020.2998816>
- W. Guo, T. Fukatsu, S. Ninomiya, Automated characterization of flowering dynamics in rice using field-acquired time-series RGB images. *Plant Methods* **11**(1), 1–15 (2015). <https://doi.org/10.1186/s13007-015-0047-9>
- L. Li, Q. Zhang, D. Huang, A review of imaging techniques for plant phenotyping. *Sensors* **14**(11), 20078–20111 (2014). <https://doi.org/10.3390/s141120078>
- J.A. Berni, P.J. Zarco-Tejada, L. Suárez, E. Fereres, Thermal and narrowband multispectral remote sensing for vegetation monitoring from an unmanned aerial vehicle. *IEEE Trans. Geosci. Remote Sens.* **47**(3), 722–738 (2009). <https://doi.org/10.1109/tgrs.2008.2010457>
- V. Sagan, M. Maimaitijiang, P. Sidike, K. Ebliimit, K.T. Peterson, S. Hartling et al., UAV-based high resolution thermal imaging for vegetation monitoring, and plant phenotyping using ICI 8640 P, FLIR Vue Pro R 640, and thermomap cameras. *Remote Sensing* **11**(3), 330 (2019). <https://doi.org/10.3390/rs11030330>
- M. Jansen, F. Gilmer, B. Biskup, K.A. Nagel, U. Rascher, A. Fischbach et al., Simultaneous phenotyping of leaf growth and chlorophyll fluorescence via GROWSCREEN FLUORO allows detection of stress tolerance in *Arabidopsis thaliana* and other rosette plants. *Funct. Plant Biol.* **36**(11), 902–914 (2009). <https://doi.org/10.1071/fp09095>
- H.M. Kalaji, A. Jajoo, A. Oukarroum, M. Brestic, M. Zivcak, I.A. Samborska et al., Chlorophyll a fluorescence as a tool to monitor physiological status of plants under abiotic stress conditions. *Acta Physiol. Plant.* **38**(4), 1–11 (2016). <https://doi.org/10.1007/s11738-016-2113-y>
- R. Zhou, B. Hyldgaard, X. Yu, E. Rosenqvist, R.M. Ugarte, S. Yu et al., Phenotyping of faba beans (*Vicia faba* L.) under cold and heat stresses using chlorophyll fluorescence. *Euphytica* **214**(4), 1–13 (2018). <https://doi.org/10.1007/s10681-018-2154-y>
- M. Zaman-Allah, O. Vergara, J. Araus, A. Tarekegne, C. Magorokosho, P. Zarco-Tejada et al., Unmanned aerial platform-based multi-spectral imaging for field phenotyping of maize. *Plant Methods* **11**(1), 1–10 (2015). <https://doi.org/10.1186/s13007-015-0078-2>
- J. Geipel, J. Link, J.A. Wirwahn, W. Claupein, A programmable aerial multispectral camera system for in-season crop biomass and nitrogen content estimation. *Agriculture* **6**(1), 4 (2016). <https://doi.org/10.3390/agriculture6010004>
- S. Liu, F. Baret, M. Abichou, F. Boudon, S. Thomas, K. Zhao et al., Estimating wheat green area index from ground-based LiDAR measurement using a 3D canopy structure model. *Agric. For. Meteorol.* **247**, 12–20 (2017). <https://doi.org/10.1016/j.agrformet.2017.07.007>
- Q. Guo, F. Wu, S. Pang, X. Zhao, L. Chen, J. Liu et al., Crop 3D—a LiDAR based platform for 3D high-throughput crop phenotyping. *Sci. Chin. Life Sci.* **61**(3), 328–339 (2018). <https://doi.org/10.1007/s11427-017-9056-0>
- M.Á. Castillo-Martínez, F.J. Gallegos-Funes, B.E. Carvajal-Gámez, G. Urriolagoitia-Sosa, A.J. Rosales-Silva, Color index based thresholding method for background and foreground segmentation of plant images. *Comput. Electron. Agric.* **178**, 105783 (2020). <https://doi.org/10.1016/j.compag.2020.105783>
- B. Romera-Paredes, P.H.S. Torr, *Recurrent instance segmentation European conference on computer vision* (Springer, Cham, 2016), pp.312–329
- N. Al-Garaawi, R. Ebsim, A.F. Alharan, M.H. Yap, Diabetic foot ulcer classification using mapped binary patterns and convolutional neural networks. *Comput. Biol. Med.* **140**, 105055 (2022). <https://doi.org/10.1016/j.combiomed.2021.105055>
- P.K. Jayapal, E. Park, M.A. Faqeerzada, Y.-S. Kim, H. Kim, I. Baek et al., Analysis of RGB plant images to identify root rot disease in Korean Ginseng plants using deep learning. *Appl. Sci.* **12**(5), 2489 (2022). <https://doi.org/10.3390/app12052489>
- X.-F. Han, H. Laga, M. Bennamoun, Image-based 3D object reconstruction: state-of-the-art and trends in the deep learning era. *IEEE Trans. Pattern Anal. Mach. Intell.* **43**(5), 1578–1604 (2019). <https://doi.org/10.1109/tpami.2019.2954885>
- G. Lin, Y. Tang, X. Zou, C. Wang, Three-dimensional reconstruction of guava fruits and branches using instance segmentation and geometry analysis. *Comput. Electron. Agric.* **184**, 106107 (2021). <https://doi.org/10.1016/j.compag.2021.106107>
- S. Yang, L. Zheng, W. Gao, B. Wang, X. Hao, J. Mi et al., An efficient processing approach for colored point cloud-based high-throughput seedling phenotyping. *Remote Sensing* **12**(10), 1540 (2020). <https://doi.org/10.3390/rs12101540>
- H. Qiaoling, X. Cui Shuqiang, Z. Y. Shanshan, Z. Yandong, Construction of the automatic quantification system for the phenotype of *Amygdalus mira* seeds based on HSV space and fitting ellipse. *Trans. Chin. Soc. Agric. Eng. (Transactions of the CSAE)* **37**(20), 202–210 (2021). <https://doi.org/10.11975/j.issn.1002-6819.2021.20.023>
- X. Zhang, Y. Zhu, Y. Su, B. Xie, Q. Gu, K. Zheng, Quantitative extraction and analysis of pear fruit spot phenotypes based on image recognition. *Comput. Electron. Agric.* **190**, 106474 (2021). <https://doi.org/10.1016/j.compag.2021.106474>
- M. Zhang, J. Zhou, K.A. Sudduth, N.R. Kitchen, Estimation of maize yield and effects of variable-rate nitrogen application using UAV-based RGB imagery. *Biosyst. Eng.* **189**, 24–35 (2020). <https://doi.org/10.1016/j.biosystemseng.2019.11.001>
- P. Song, H. Zhang, C. Wang, B. Luo, W. Lu, P. Hou, Design and experiment of high throughput automatic measuring device for

- corn. Trans. Chin. Soc. Agric. Eng. **33**(16), 41–47 (2017). <https://doi.org/10.1109/icae.2011.5943742>
26. Ren, M., & Zemel, R. S.: End-to-end instance segmentation with recurrent attention. Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 6656–6664. (2017) <https://doi.org/10.1109/cvpr.2017.39>
27. Xu, L., Li, Y., Sun, Y., Song, L., & Jin, S.: Leaf instance segmentation and counting based on deep object detection and segmentation networks. 2018 Joint 10th International Conference on Soft Computing and Intelligent Systems (SCIS) and 19th International Symposium on Advanced Intelligent Systems (ISIS). IEEE, pp. 180–185. (2018) <https://doi.org/10.1109/scis-isis.2018.00038>
28. C. Wang, P. Du, H. Wu, J. Li, C. Zhao, H. Zhu, A cucumber leaf disease severity classification method based on the fusion of DeepLabV3+ and U-Net. Comput. Electron. Agric. **189**, 106373 (2021). <https://doi.org/10.1016/j.compag.2021.106373>
29. Z. Wu, R. Yang, F. Gao, W. Wang, L. Fu, R. Li, Segmentation of abnormal leaves of hydroponic lettuce based on DeepLabV3+ for robotic sorting. Comput. Electron. Agric. **190**, 106443 (2021). <https://doi.org/10.1016/j.compag.2021.106443>
30. Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H., Encoder-decoder with atrous separable convolution for semantic image segmentation. Proceedings of the European conference on computer vision (ECCV). pp. 801–818. (2018) https://doi.org/10.1007/978-3-030-01234-2_49
31. N. Altini, G.D. Cascarano, A. Brunetti, F. Marino, M.T. Rocchetti, S. Matino et al., Semantic segmentation framework for glomeruli detection and classification in kidney histological sections. Electronics **9**(3), 503 (2020). <https://doi.org/10.3390/electronics9030503>
32. Z. Song, Z. Zhou, W. Wang, F. Gao, L. Fu, R. Li et al., Canopy segmentation and wire reconstruction for kiwifruit robotic harvesting. Comput. Electron. Agric. **181**, 105933 (2021). <https://doi.org/10.1016/j.compag.2020.105933>
33. Y. Sun, Y. Yang, G. Yao, F. Wei, M. Wong, Autonomous crack and bughole detection for concrete surface image based on deep learning. IEEE Access **9**, 85709–85720 (2021). <https://doi.org/10.1109/access.2021.3088292>
34. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C., Mobilenetv2: inverted residuals and linear bottlenecks. Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4510–4520. (2018) <https://doi.org/10.1109/cvpr.2018.00474>
35. M. Pourshamsi, J. Xia, N. Yokoya, M. Garcia, M. Lavalle, E. Potier et al., Tropical forest canopy height estimation from combined polarimetric SAR and LiDAR using machine-learning. ISPRS J. Photogramm. Remote. Sens. **172**, 79–94 (2021). <https://doi.org/10.1016/j.isprsjprs.2020.11.008>
36. J. Zhang, T. Cheng, W. Guo, X. Xu, H. Qiao, Y. Xie et al., Leaf area index estimation model for UAV image hyperspectral data based on wavelength variable selection and machine learning methods. Plant Methods **17**(1), 1–14 (2021). <https://doi.org/10.21203/rs.3.rs-131883/v1>
37. Y. Fu, G. Yang, X. Song, Z. Li, X. Xu, H. Feng et al., Improved estimation of winter wheat aboveground biomass using multiscale textures extracted from UAV-based digital images and hyperspectral feature analysis. Remote Sensing **13**(4), 581 (2021). <https://doi.org/10.3390/rs13040581>
38. Y. Guo, Y. Fu, F. Hao, X. Zhang, W. Wu, X. Jin et al., Integrated phenology and climate in rice yields prediction using machine learning methods. Ecol. Ind. **120**, 106935 (2021). <https://doi.org/10.1016/j.ecolind.2020.106935>
39. J. Su, D. Yi, B. Su, Z. Mi, C. Liu, X. Hu et al., Aerial visual perception in smart farming: field study of wheat yellow rust monitoring. IEEE Trans. Industr. Inf. **17**(3), 2242–2249 (2020). <https://doi.org/10.1109/tii.2020.2979237>
40. M. Yoosefzadeh-Najafabadi, H.J. Earl, D. Tulpan, J. Sulik, M. Eskandari, Application of machine learning algorithms in plant breeding: predicting yield from hyperspectral reflectance in soybean. Front. Plant Sci. **11**, 2169 (2021). <https://doi.org/10.3389/fpls.2020.624273>
41. Raja, P., Olenskyj, A., Kamangir, H., & Earles, M., Simultaneously Predicting Multiple Plant Traits from Multiple Sensors via Deformable CNN Regression. arXiv preprint [arXiv:2112.03205](https://arxiv.org/abs/2112.03205), (2021). <https://doi.org/10.48550/arXiv.2112.03205>
42. Bloice, M. D., Stocker, C., & Holzinger, A., Augmentor: an image augmentation library for machine learning. arXiv preprint [arXiv:1708.04680](https://arxiv.org/abs/1708.04680), (2017). <https://doi.org/10.48550/arXiv.1708.04680>
43. Chollet, F., Xception: Deep learning with depthwise separable convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1251–1258 (2017). <https://doi.org/10.1109/cvpr.2017.195>
44. A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan et al., Pytorch: an imperative style, high-performance deep learning library. Adv. Neural Inform. Process. Syst. (2019). <https://doi.org/10.48550/arXiv.1912.01703>
45. D. Riehle, D. Reiser, H.W. Griepentrog, Robust index-based semantic plant/background segmentation for RGB-images. Comput. Electron. Agric. **169**, 105201 (2020). <https://doi.org/10.1016/j.compag.2019.105201>
46. C. Rother, V. Kolmogorov, A. Blake, “GrabCut” interactive foreground extraction using iterated graph cuts. ACM Trans. Graph. (TOG) **23**(3), 309–314 (2004). <https://doi.org/10.1145/1186562.1015720>
47. D. Wu, C. Zhang, L. Ji, R. Ran, H. Wu, Y. Xu, Forest fire recognition based on feature extraction from multi-view images. Traitement du Signal (2021). <https://doi.org/10.18280/ts.380324>
48. Ronneberger, O., Fischer, P., & Brox, T., U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention. Springer, pp. 234–241. (2015) <https://doi.org/10.48550/arXiv.1505.04597>
49. Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J., Pyramid scene parsing network. Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2881–2890. (2017) <https://doi.org/10.48550/arXiv.1612.01105>
50. W. Liu, Y. Li, J. Liu, J. Jiang, Estimation of plant height and aboveground biomass of *Toona sinensis* under drought stress using RGB-D imaging. Forests **12**(12), 1747 (2021). <https://doi.org/10.3390/f12121747>
51. M. Maimaitijiang, V. Sagan, P. Sidike, M. Maimaitiyiming, S. Hartling, K.T. Peterson et al., Vegetation index weighted canopy volume model (CVMVI) for soybean biomass estimation from unmanned aerial system-based RGB imagery. ISPRS J. Photogramm. Remote. Sens. **151**, 27–41 (2019). <https://doi.org/10.1016/j.isprsjprs.2019.03.003>
52. W. Mao, Y. Wang, Y. Wang, Real-time detection of between-row weeds using machine vision. 2003 ASAE Annual Meeting. Am. Soc. Agric. Biol. Eng. (2003). <https://doi.org/10.13031/2013.15381>
53. J. Torres-Sánchez, J.M. Pena, A.I. de Castro, F. López-Granados, Multi-temporal mapping of the vegetation fraction in early-season wheat fields using images from UAV. Comput. Electron. Agric. **103**, 104–113 (2014). <https://doi.org/10.1016/j.compag.2014.02.009>
54. D.M. Woebbecke, G.E. Meyer, K. Von Bargen, D.A. Mortensen, Plant species identification, size, and enumeration using machine vision techniques on near-binary images. Optics Agric. For. (1993). <https://doi.org/10.1117/12.144030>
55. J. Bendig, K. Yu, H. Aasen, A. Bolten, S. Bennertz, J. Broscheit et al., Combining UAV-based plant height from crop surface models, visible, and near infrared vegetation indices for biomass

- monitoring in barley. *Int. J. Appl. Earth Obs. Geoinf.* **39**, 79–87 (2015). <https://doi.org/10.1016/j.jag.2015.02.012>
56. X. Zhou, H. Zheng, X. Xu, J. He, X. Ge, X. Yao et al., Predicting grain yield in rice using multi-temporal vegetation indices from UAV-based multispectral and digital imagery. *ISPRS J. Photogramm. Remote. Sens.* **130**, 246–255 (2017). <https://doi.org/10.1016/j.isprsjprs.2017.05.003>
57. J. Du, B. Li, X. Lu, X. Yang, X. Guo, C. Zhao, Quantitative phenotyping and evaluation for lettuce leaves of multiple semantic components. *Plant Methods* **18**(1), 1–18 (2022). <https://doi.org/10.1186/s13007-022-00890-2>
58. J. Du, X. Lu, J. Fan, Y. Qin, X. Yang, X. Guo, Image-based high-throughput detection and phenotype evaluation method for multiple lettuce varieties. *Front. Plant Sci.* **11**, 563386 (2020). <https://doi.org/10.3389/fpls.2020.563386>
59. J.P. Goncalves, F.A. Pinto, D.M. Queiroz, F.M. Villar, J.G. Barbedo, E.M. Del Ponte, Deep learning architectures for semantic segmentation and automatic estimation of severity of foliar symptoms caused by diseases or pests. *Biosys. Eng.* **210**, 129–142 (2021). <https://doi.org/10.1016/j.biosystemseng.2021.08.011>
60. S. Zhang, Z. Wang, Z. Wang, Method for image segmentation of cucumber disease leaves based on multi-scale fusion convolutional neural networks. *Trans. Chin. Soc. Agric. Eng.* **36**(6), 149–157 (2020)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.